DOCUMENT RESUME

ED 135 802 TM 005 733

AUTHOR Arneklev, Bruce; And Others
TITLE Optimum Range of Difficulty for Linking Items.
PUB DATE [Apr 76]
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Elementary School Mathematics; Goodness of Fit; *Item Analysis; *Measurement Techniques; *Scoring Formulas; Tests
IDENTIFIERS *Rasch Model; *Test Linking

ABSTRACT

One of the most important contentions of the Rasch model of item analysis is that two tests of the same trait, having some items in common, can be linked together using a "linking constant" derived from the common items. This would be accomplished by administering both tests to a sample of testees, calibrating the items of the tests independently, and then linking the test by a linking constant which would be the difference between the average calibration of the common items on test one and the independently determined average calibration of the common items on test two. An important question regarding this linking procedure is whether the accuracy of the link is affected significantly when the items used for the linking represent a restricted range of difficulty relative to all items in the tests being linked. This paper empirically addresses this question with a variety of analyses using test results from mathematics tests administered to samples of fourth through seventh grade pupils. While the analyses varied with respect to the specific questions that they addressed, the overall conclusion was that for these tests, restriction in range of item difficultly did not present serious difficulties for test linking using the Rasch procedures. (Author/JKS)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED135802

April 9, 1976

12.13

OPTIMUM RANGE OF DIFFICULTY FOR LINKING ITEMS

Bruce Arneklev, Tacoma Public Schools
Donald Gee, Multnomah County IED
George Ingebo, Portland Public Schools

Problem.

Does restricting the range of difficulty of linking items affect the linking constant between two tests?

Introduction.

One of the important assumptions of the Rasch scaling procedure is that by giving two different tests, with each test having some items in common with the other test, calibrated item difficulty values for the linking items in both tests can be derived from performance of students on those items, and these values can be adjusted to the same scale. This is done through a "linking constant" which is the difference between the average of calibrated item difficulties for the linking items in one test and the average for those same items in the test sharing those linking items. The calibrated item difficulty values for the common(linking) items will normally differ in the two tests because the performance level of students taking the two tests will usually be different. An important question regarding this linking procedure is whether the accuracy of the link is affected significantly if the items used for linking represent a restricted range of difficulty relative to all items in the test being linked. If the accuracy of the link is not affected, the number and difficulty range of items used for linking purposes can be reduced, thus simplifying the development of linked tests.

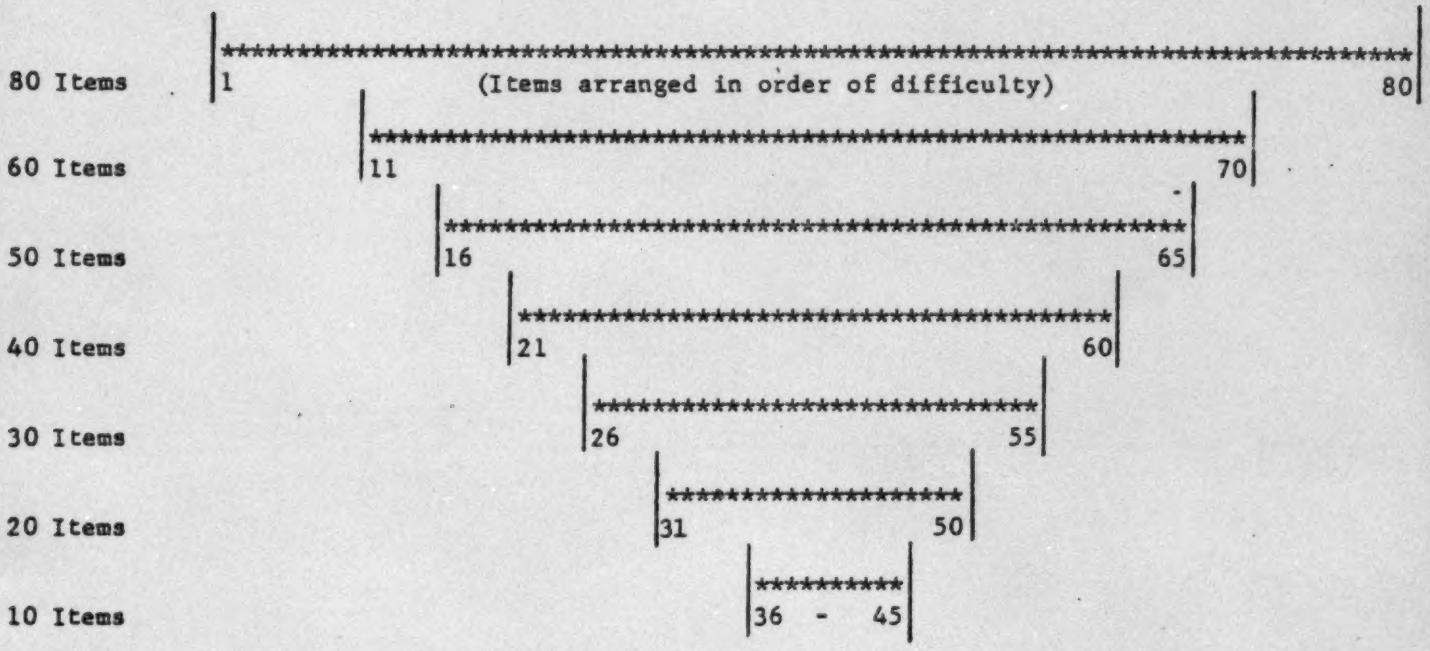
TM005 733

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Procedure I.

I. An eighty-item test of fourth grade mathematics was administered in the Fall of 1974. Using the test scores of approximately 400 students, the items were rank ordered in level of difficulty using their Rasch difficulty calibrations.



The groups of 60, 50, 40, 30, 20 and 10 items were calibrated separately by the Rasch technique. It is important to note that the sub-groups of items that were recalibrated were not readministered, but that the item difficulty data obtained from the Fall 1974 administration were used both in calibrating the eighty-item pool and in calibrating each subset of items as described above.

All seven tests from lengths 80 to 10 were linked to the 80-and 60-item tests using all of the items in common for each linkage.

The linking constant needed to put the 60-item and the 80-item tests on the same scale was found to be .388.

The linking constants derived by using the 50, 40, 30, 20, and 10 items of increasingly restricted range of difficulty to link the 60-and 80-item tests are compared in the table below. These constants are expressed in terms of Rasch scaled score units. These values are obtained by multiplying the Rasch difficulty calibration by 4.53 and adding the constant of 50.

Direct Link	. . . . .	Link 60 to 80 = .388
Indirect Links	60 to 50 (- .083) + 50 to 80 (.469)	= .386
	60 to 40 (- .189) + 40 to 80 (.574)	= .385
	60 to 30 (- .394) + 30 to 80 (.777)	= .383
	60 to 20 (- .734) + 20 to 80 (1.113)	= .379
	60 to 10 (- .865) + 10 to 80 (1.243)	= .378

The change in magnitude of linking factors of a .01 is negligible, much less than the standard error of a single item close to the middle range of difficulty.

Procedure II.

The procedures were replicated on a fourth-grade test given Spring of 1975 with the difference being that the short tests were selected to be ranked 10 items more difficult than the short Fall tests were.

A sixty-item test was selected by excluding the 20 easiest items. Successive short test selections were made by dropping five items from each

end of the previous selection.

Again, the 60, 50, 40, 30, 20 and 10 item tests were calibrated separately and all seven tests linked to the 60 and 80 item tests. A table of linking constants is shown below.

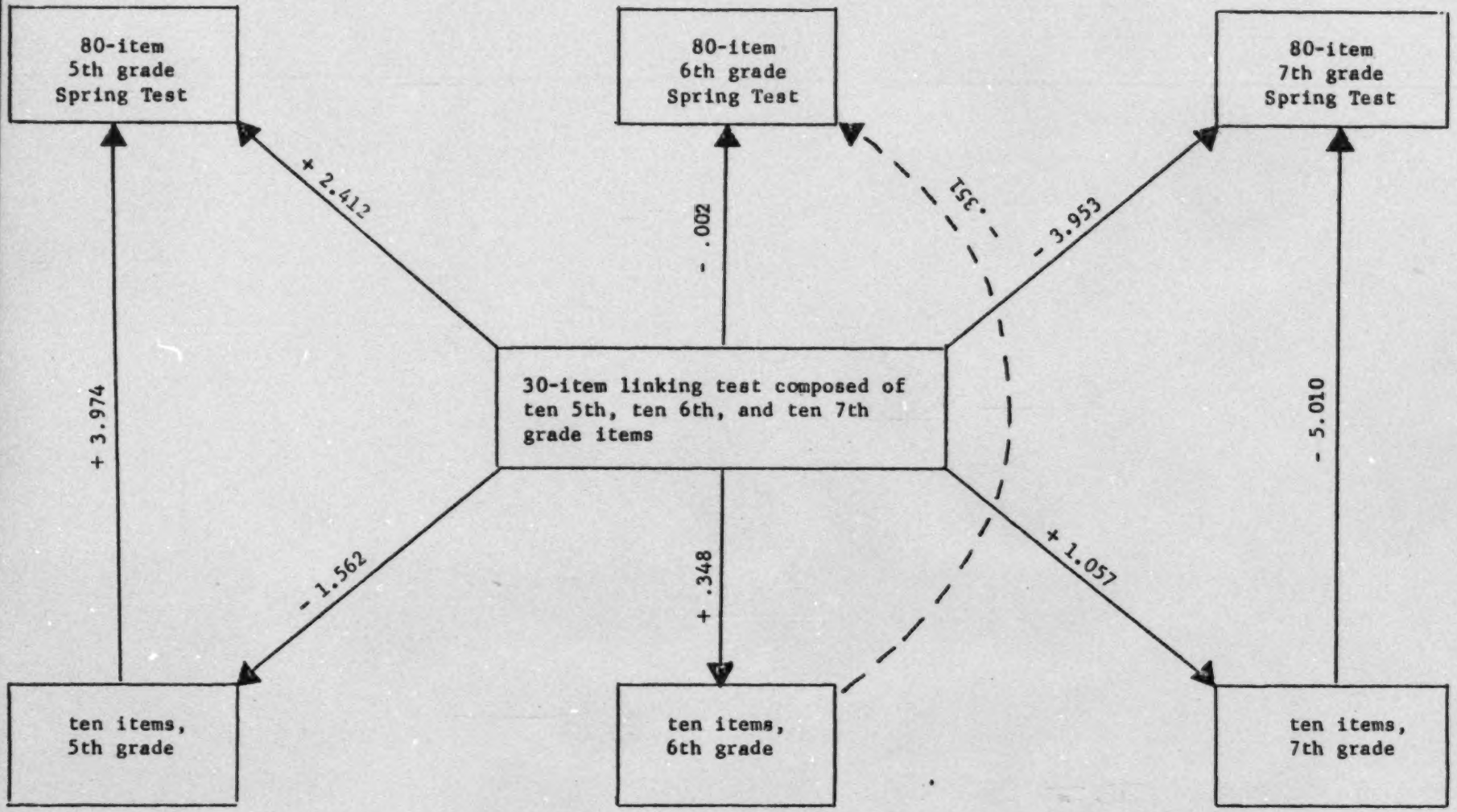
Direct Link	Link 60 to 80	= 2.456
Indirect Links	60 to 50 (.041) + 50 to 80 (2.411)	= 2.452
	50 to 40 (.120) + 40 to 80 (2.331)	= 2.451
	40 to 30 (.216) + 30 to 80 (2.237)	= 2.453
	30 to 20 (.151) + 20 to 80 (2.297)	= 2.448
	20 to 10 (.160) + 10 to 80 (2.288)	= 2.448

Procedure III.

The effects of restricted range of linking items was further examined by using ten hard items from a fifth grade test, ten middle-difficulty items from a sixth grade test and ten easy items from a seventh grade test to construct a 30-item linking test. Each group of ten items was calibrated separately, was calibrated in the 30-item linking test, and was calibrated in an 80-item test. In each case the same items were involved but calibrated in groups of items of varying ranges of difficulty. This linking experiment was carried out in the Spring and again in the Fall as shown in the tables and diagrams that follow.

Spring Results.	Linking Constant
10 5th grade items to 80-item 5th grade test	= 3.974
10 5th grade items to 30-item linking test (1.562) and 30-item linking test to 80-item 5th grade test (2.412)	= 3.974
- - -

Figure 1.


80-item
5th grade
Spring Test
80-item
6th grade
Spring Test
80-item
7th grade
Spring Test
+ 2.412
- .002
.351
- 3.953
+ 3.974
30-item linking test composed of
ten 5th, ten 6th, and ten 7th
grade items
- 5.010
- 1.562
+ .348
+ 1.057
ten items,
5th grade
ten items,
6th grade
ten items,
7th grade

Fall Results.

10 5th grade items to 80-item 5th grade test = 4.164
10 5th grade items to 30-item linking test (1.562) and
30-item linking test to 80-item 5th grade test (2.603) = 4.164

- -

10 6th grade items to 80-item 6th grade test = - .213

10 6th grade items to 30-item linking test (-.348) and
30-item linking test to 80-item 6th grade test (.135) = - .350

- -

10 7th grade items to 80-item 7th grade test = - 5.006

10 7th grade items to 30-item linking test (-1.057) and
30-item linking test to 80-item 7th grade test (-3.949) = - 5.006

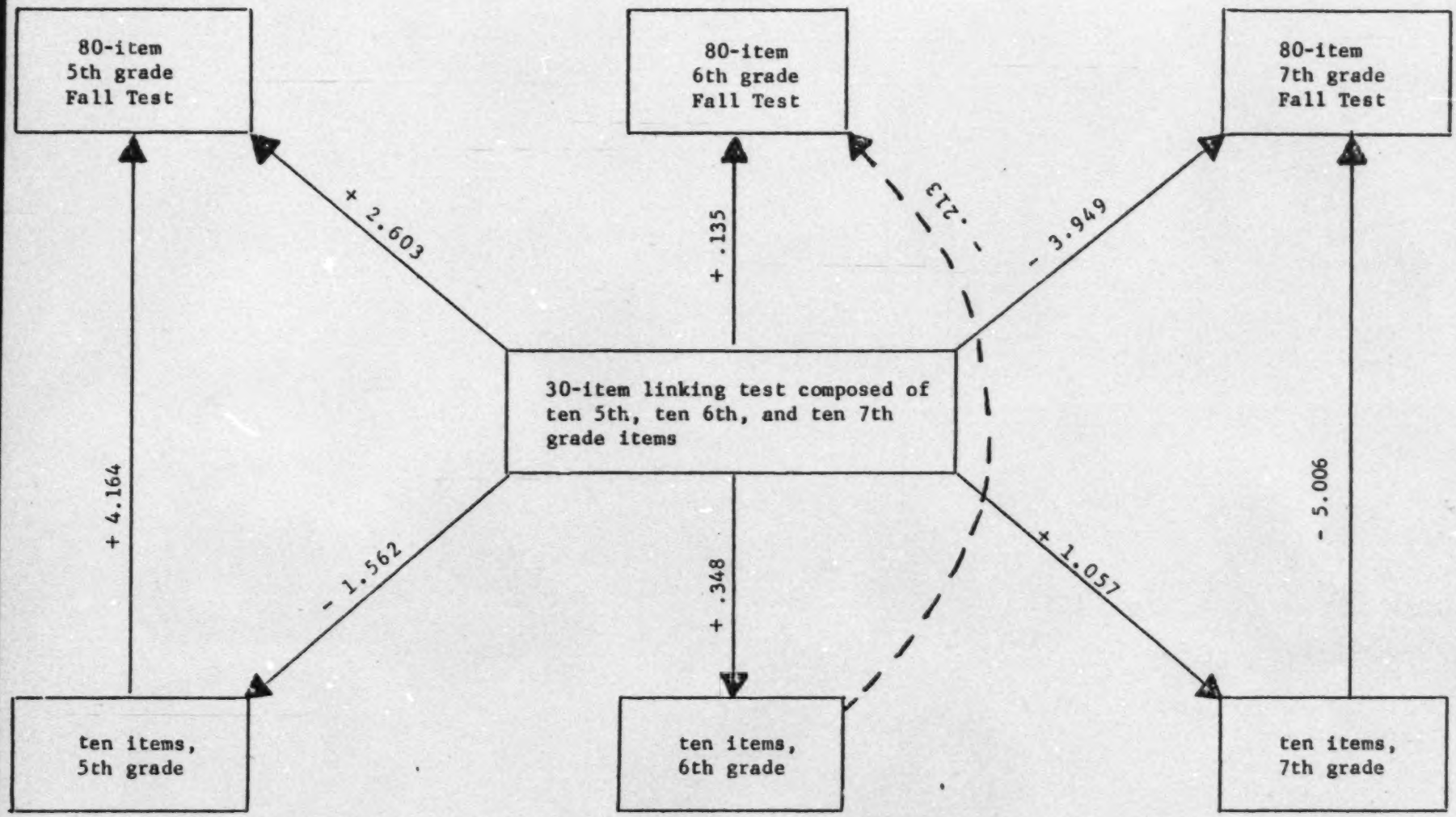
- -

80-item 5th grade test to 30-item linking test (-2.603) and
30-item linking test to 80-item 7th grade test (-3.949) = - 6.552

80-item 5th grade test to 10 5th grade items (-4.164) and
10 5th grade items to 30-item linking test (1.057) and
30-item linking test to 10 7th grade items (1.057) and
10 7th grade items to 80-item 7th grade test (-5.006) = - 6.551

- -

Figure 2.


80-item
5th grade
Fall Test
80-item
6th grade
Fall Test
80-item
7th grade
Fall Test
+ 2.603
+ .135
.213
- 3.949
30-item linking test composed of
ten 5th, ten 6th, and ten 7th
grade items
+ 4.164
- 1.562
+ .348
+ 1.057
- 5.006
ten items,
5th grade
ten items,
6th grade
ten items,
7th grade

Procedure IV.

Linking through use of items severely restricted in range of difficulty was tried by overlapping seven mathematics tests of selected difficulty ranges. Going from very easy to very difficult for seventh grade students, the tests were five Rasch scaled scores in range of difficulty with an overlap of two scaled scores with the test more difficult and with the one less difficult. Since the easiest level was only administered to 50 students, the link between it and the next easiest test was not included in the investigation. The five links with a range of only two Rasch-scaled scores each were compared with the results of linking the six tests to more extended range links with a third set of calibrated items. Two different full-length survey tests were used and the average of the linking constants was considered the best linking estimate. The comparisons of the five restricted-range links with the best estimated link from wide-range links are tabled below.

Linked Tests	Best Estimate	Restricted Range	Difference
2 - 3	1.90	1.82	.08
3 - 4	2.86	2.95	- .09
4 - 5	2.76	2.77	- .01
5 - 6	2.65	2.42	.23
6 - 7	2.92	2.76	.16
2 through 7	13.09	12.72	.37

Note that this procedure involves an important difference compared to procedures I, II and III. Subsets of items were recalibrated under procedures I, II and III from the single administration with the subsets

drawn from the total set of 80 being recalibrated on the basis of the performance on the 80 items were administered. In procedure IV, subtests were readministered to different groups of students before recalibration. Conditions under procedure IV constitute a more severe test and approximate the results one might expect if a pool of calibrated items was used from which to select a subset as a new test.

Conclusion.

Severely restricting the range of linking items does not adversely affect linking mathematics tests at approximately seventh-grade level.